立体的な音場を実現する革新的な音楽体験「360 Reality Audio」

ソニーグループ株式会社
R&Dセンター
Tokyo Laboratory 20
知念（ちねん）　徹（とおる）

ソニー株式会社
HES事業本部
ライセンスビジネス推進室
海鋒（かいほこ）　俊哉（としや）

1. はじめに―「360 Reality Audio」概要

ソニーは、今年1月に開催されたテクノロジーイベント「CES 2021」にて「360 Reality Audio」の新たな取組みを発表した。この「360 Reality Audio」は、ソニーの360立体音響技術を使った、没入感のある立体的な音場を実現する音楽体験であり、「サンロクマル・リアリティオーディオ」と読む。2019年より、欧米などで複数のストリーミングサービスから対応楽曲の配信を開始している。

「CES2021」では、今後の、立体的な音場に映像を組み合わせた臨場感のあるビデオコンテンツの配信や、対応楽曲制作ツール「360 Reality Audio Creative Suite」の共同開発、対応機器の拡大、360 Reality Audioに関連する独自技術の他社へのライセンス提供を進めることを発表した。

また、「CES 2021」期間中には、360 Reality Audio初のビデオコンテンツとして、ザラ・ラーソンによるパフォーマンスの配信も行われ、ご好評をいただいた。

■図1.「CES2021」期間中に配信されたパフォーマンス配信

その後、2021年3月には、日本国内への本格的導入も発表を行った。既に邦楽アーティストの楽曲を含めた豊富なコンテンツの配信が開始されており、多くのお客さまに、新しい音楽体験をお楽しみいただいている。

「360 Reality Audio」では、オブジェクトベースの360立体音響技術を活用しており、クリエイターが、360度に広がる全球に、ボーカルやコーラス、楽器などの音源一つひとつに位置情報を付けて自由に配置し、それらを動かすことが可能で、その創造性のままに自由な音楽表現が可能となる。再生時には、リスナーの皆さまに、まるでアーティストの生演奏に囲まれているかのような、没入感のある立体的な音場を体感いただくことができる。

本稿では、この「360 Reality Audio」を支える、オブジェクトベースの配信メディアフォーマットと、コンテンツワークフロー及びそれらに活用されている技術を紹介する。

2.「360 Reality Audio」配信メディアフォーマット

近年、360度映像サービスやバーチャルリアリティ（VR）製品が急速に普及しているが、それらに利用されるオーディオ技術も、新たな技術が開発されてきた。いままでの2チャンネルステレオや、5.1チャンネルサラウンドなどでは、水平方向のみの音の広がりだったものが、それに加えて、高さ方向の音の広がりも持つ立体的な音場を実現する技術や、音源の音像を正確に立体的な位置で表現するためのオブジェクトベースオーディオ技術などが使われ始めている。

ところで、これらに関連する技術は、国際技術規格「MPEG-H 3D Audio」としてISO/IEC 23008-3でも規格化されており、そのコアコーデックは、高効率音声圧縮が可能なUSAC（Unified Speech and Audio Coding, ISO/IEC 14496-3）が指定されている。また、立体空間表現のために、オーディオトラックとして、オブジェクトベース、チャンネルベース、HOA（高次アンビソニックス）が規定されている。さらに、それらの対応の有無や、そのオーディオトラックの上限数な

どの組合せが、それぞれ「Profile」及び「Level」として規定されており、Low Complexity Profileや、2020年に新たに規格化された、Baseline Profileなどがある。

「360 Reality Audio」では、世界中の多くの皆さまに立体音響コンテンツをお楽しみいただけるよう、この、国際技術規格であるMPEG-H 3D Audioをベースとした配信メディアフォーマットを規定した。

この配信メディアフォーマットを策定する際には、(1) クリエイターの創造性のままに自由な音楽表現ができ、その想いを十分に伝えることができること、そして、(2) 多くのオーディオ機器やスマートフォンアプリで、容易に、高音質で高い臨場感を再生できること、の両面を考慮した。具体的には、ソニーグループのアーティストとサウンドエンジニア、オーディオ機器やスマートフォンの技術者、技術規格開発担当者が協力し、収録実験と検討を行い、それを基に、(1) (2) を両立できる配信メディアフォーマットの技術規格化の議論をした。

その結果、オブジェクトベースのオーディオトラックのみで、十分に高音質で臨場感のある音楽配信ができることが検証できた。また、必要なトラック数や、そのビットレートについても検討された。これらの結果を基に、「360 Reality Audio」の配信メディアフォーマットでは、オブジェクトベースのみを規定し（**図2**）、さらに「レベル1, 2, 3」（**表**）の詳細を規定した。

■表.「360 Reality Audio」配信メディアフォーマットのレベルと、規定トラック数、トータルビットレート

Level	Max. number of total Tracks	Max. total Bitrate
1	10 Object Tracks	640kbps
2	16 Object Tracks	1,024kbps
3	24 Object Tracks	1,536kbps

3段階の「レベル」は、配信サービスのネットワークの伝送帯域や、再生側のオーディオ機器またはスマートフォンンアプリでの信号処理量に応じた再生クオリティを選択的に提供するために設けられた。配信サービスでは、それぞれの状況に応じて、自動的に最適なレベルが選択され、配信される。

なお、MPEG-H 3D Audio Baseline Profileにおいても、表に示されるようなレベルが規定されており、オブジェクトベース及びチャンネルベースが利用可能である。

他方で、「360 Reality Audio」配信メディアフォーマットでは、オブジェクトベースのみが使われる。

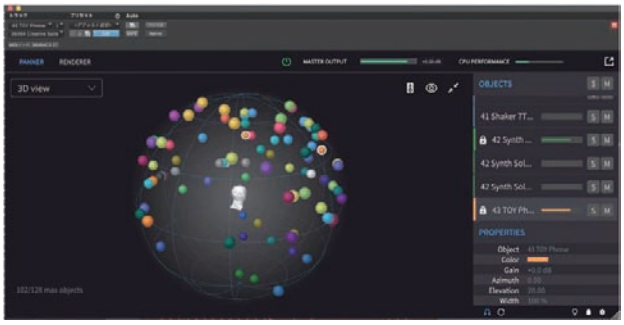
3.「360 Reality Audio」コンテンツワークフロー

図2は、「360 Reality Audio」の、コンテンツの流れを示すコンテンツワークフローである。

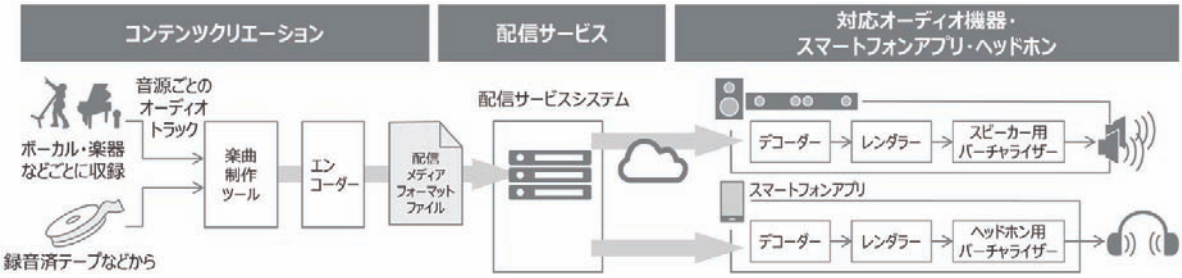
コンテンツクリエーションでは、まず、音源ごとのオーディオトラック、すなわち、ボーカルやコーラス、楽器など、音源ごとのデータが用意される。これは、「360 Reality Audio」向けに新たに収録されたものだけでなく、過去に収録された録音テープなどからも用意することができる。

次に、楽曲制作ツールにて、それぞれの音源を、360度に拡がる全球音場内に配置する。これらを、時間軸で、自由に動かすことも可能である。リスナーから見て、その前方や上方、後方に、ボーカル、コーラス、楽器を配置し、さらに、ボーカルの動きも表現することができる。

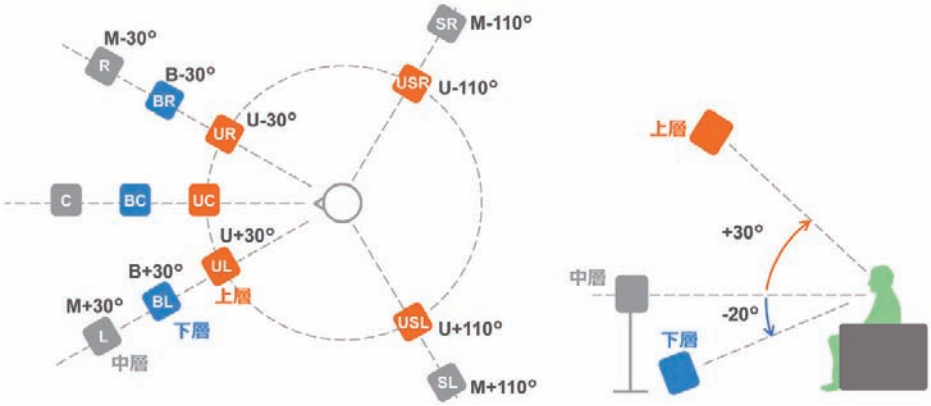
図3に、「CES2021」で発表した、対応楽曲制作ツール「360 Reality Audio Creative Suite」での画面一例を示す。色



■図3.「360 Reality Audio」楽曲制作ツール画面の一例



■図2.「360 Reality Audio」コンテンツワークフロー概要



■図4.「360 Reality Audio」のスピーカー配置

のついたボールが一つひとつの音源を示している。

「360 Reality Audio」では、音場内に音源位置を再現する際のスピーカー配置として、図4に示す、合計13個のスピーカーの配置を定めている。制作スタジオでは、このスピーカー配置のシステムを使い、全天球音場での音源配置の音響確認が行われている。

音響確認のあと、楽曲制作ツールを使い、各音源のオーディオトラックに付随するかたちで、その位置情報データを生成する。

次に、これらを「360 Reality Audio」のエンコーダーに入力し、「360 Reality Audio」配信メディアフォーマットに準じたエンコードを行う。図5にエンコーダーとデコーダーの中核部分のブロックダイアグラム概要図を示す。

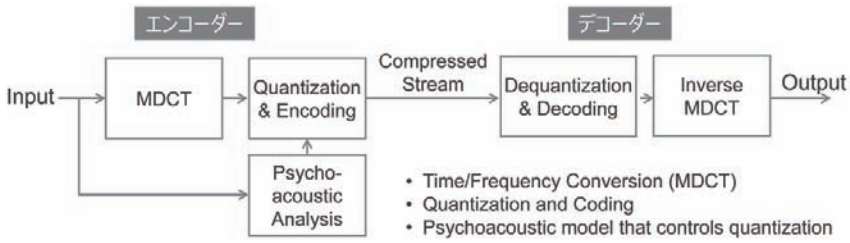
中核となるエンコーダーは、各オーディオトラック（以下、オブジェクト）を圧縮する。各オブジェクトは、まずMDCT（Modified Discrete Cosine Transform）により時間から周波数の領域に変換される。次に、変換された係数は、人間の聴覚心理モデル（Psychoacoustic model）を用いて周波数ごとで異なる量子化が行われる。最後に、量子化データは、「360 Reality Audio」配信メディアフォーマットに準じて符号化される。対応するデコーダーは、エンコーダーの逆処理を行っているので、ここでは説明を割愛する。

表に示したとおり、「360 Reality Audio」では、最大24個のオブジェクトが使われるが、エンコードの際には、それら各オブジェクトのエンコード後のビットレート配分が重要になる。例えば、オブジェクトのうち、メインボーカルの配分を優先するのはもちろんであるが、前方下層の3個のスピーカー部分に配置される音源も、再生時の空間認知面から優先すべきオブジェクトになり、また、それ以外のボーカルやコーラス、楽器などのバランスも考慮して、ビットレート配分を決める必要がある。

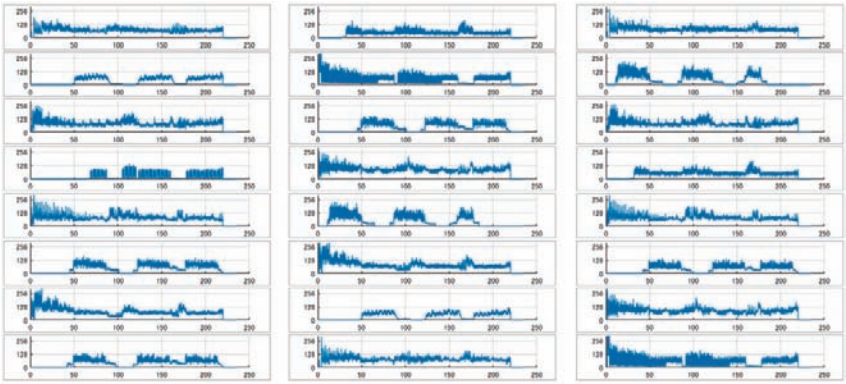
ソニーが開発した、「360 Reality Audio」エンコーダーでは、これらのオブジェクトごとのビットレートを、時間軸方向でも細かく調整し、全体最適化を行いながらダイナミックに配分している。これにより、トータルビットレートを抑えながらも、楽曲再生時に高い臨場感が得られるクオリティを実現している。

図6に、24個のオブジェクトを圧縮する際の例を示す。各オブジェクトの時間に対するビットレートが示されている。図6においてビットレートは最大288kbit/secまで増加しており、時間軸でダイナミックにコントロールされていることが分かる。

このようにして作成された「360 Reality Audio」配信メディアフォーマットの楽曲ファイルは、配信サービス側に送られ、そのサーバーシステムから、オーディオ機器やスマートフォン



■図5.「360 Reality Audio」エンコーダー・デコーダーブロックダイアグラム概要



例) 24オブジェクトオーディオトラック、レベル3:トータルビットレート 1.536Mbps
横軸:楽曲開始からの時間(秒)、縦軸:配分されたビットレート(kbps)

■図6.「360 Reality Audio」エンコーダーでのオブジェクトごとのビットレート配分例

アプリに向けてストリーミング配信が行われ、リスナーの持つ対応オーディオ機器やスマートフォンアプリで再生される。

再生の際には、まず、デコーダーにて各オブジェクトがデコードされ、その後、それらの位置情報データを基に、前述の13個のスピーカーに相当するオーディオ信号が生成される。この機能をレンダラーと呼ぶ。これら13個のスピーカーを持つオーディオ機器の場合には、配置位置に合わせてそれらを再生すれば、臨場感のある楽曲再生ができる。

一方、小型のオーディオ機器などで、13個のスピーカー構成ができない場合には、バーチャライザーが使われる。これは、少ないスピーカー数であっても、音響信号処理によって、あたかも13個のスピーカーが正しい位置に配置されているかのように再生する技術であり、サウンドバーやワイヤレススピーカーなどに搭載されている。

スマートフォンアプリで再生する場合には、ヘッドホン用のバーチャライザーが使われ、ステレオヘッドホンで聴いた場合でも、リスナーの周囲に13個のスピーカーが配置されているかのような音場を仮想的に生成し、臨場感のある再生を実現している。これに加え、リスナー個人個人の耳の写真から、その形状に合わせて個人最適化した音響信号処理を行うことにより、より高い音源位置精度での再現ができる技術も活用されている。

ソニーは、これら「360 Reality Audio」向けのデコーダー・レンダラー・バーチャライザーの各技術も開発し、対応オーディオ機器やスマートフォンアプリに導入している。

4. より広がる、今後の立体音響の楽しみかた

前述のように、「360 Reality Audio配信メディアフォーマットは、MPEG-H 3D Audioをベースとし、再生互換性が確保されている。また、映像配信サービスにおいても、映像と「360 Reality Audio」配信メディアフォーマットを組み合わせて、今後の対応テレビで再生することが可能である。

将来、次世代テレビ放送や配信サービスで、MPEG-H 3D Audio Baseline Profile規格が採用され、広く活用されることが期待されており、より多くのテレビやサウンドバーで立体音響による臨場感をお楽しみいただけるようになる。

■図7. 立体音響対応機器への伝送

5. おわりに

昨今、バーチャルライブやコンサートが多く開催されるようになった。お互いの場所は離れていても、クリエイターの想いと表現を、リスナーの皆さまが、その没入感、臨場感を感じながら楽しめる配信サービスが期待されている。また、今後の360度映像サービスやバーチャルリアリティ（VR）サービスが普及するなかで、立体音響空間での音像表現がますます期待されるであろう。

こうした期待にも、本稿で説明した「360 Reality Audio」の立体音響技術を活用して、お応えすることができると考える。コンテンツクリエーションの方法と、再生側での臨場感表現についても、それぞれの用途に合わせて工夫ができる。

今後、これらを含む多くの分野に、ソニーの360立体音響技術を広く活用いただけるよう、さらなる展開を検討していく。

「360 Reality Audio」について、ライセンスなどの詳細情報や問い合わせ先は、以下に掲載している：
https://www.sony.co.jp/Products/360RA/licensing/